

WO 2004/049194

PCT/KR2003/002569

Claims

1. A system for parsing a web-document based on elements, which is provided to an application of a handheld terminal when the system calls the web-document to provide it to the handheld terminal, comprising:
- 5 a word parser for separating a token on the basis of markup and non-markup by referring to a token table for all markup data necessary for kind of document to be supported; and
- 10 a syntax parser for parsing a contents model on the basis of document type definition (DTD) of each document, parsing each syntax on the basis of the result of parsing the contents model, and generating a tree-based object on the basis of graphic user interface (GUI) of the terminal.
2. The system of claim 1, wherein the word parser comprises:
- 15 a comment parser for processing a comment and a space;
- a markup start parser for recognizing a markup start tag and generating a token;
- an attribute parser for parsing an attribute and generating a token; and
- a parsed character data analyzer for analyzing parsed character data and generating a token.
- 20
3. The system of claim 1, wherein the syntax parser comprises:
- an XML verifier for verifying whether a corresponding document is composed suitable for each DTD on the basis of the token generated by the word parser; and
- 25 a terminal GUI-based object generator for matching the analyzed markup and a GUI of the terminal.
4. The system of any one of claims 1 through 3, wherein the parsing system integrally parses a web-document composed on the basis of any one of SGML and XML related to HTML, XHTML, mHTML, cHTML, WML and HDML.
- 30
5. The system of any one of claims 1 through 3, wherein the parsing system can be applied to any handheld terminal and select kind of an element to be parsed according to specification of each of the terminals.
- 35
6. A method for parsing a called web-document of a web-server, the method comprising the steps of:

WO 2004/049194

PCT/KR2003/002569

- (a) reading a token from the web-document and parsing the token;
- (b) if the token is not a defined start tag or if the token is a comment or a space as result of the step (a), ignoring the token, and when the defined start tag is read, parsing an attribute of an element from the token;
- 5 (c) parsing the attribute of the element from the token, storing GUI-related information of the element, and parsing contents of the element;
- (d) as the result of the step (c), if the contents of the element are parsed character data, storing GUI-related information of the contents, and if the contents of the element are not the parsed character data, reading data until an end tag appears; and
- 10 (e) in case the contents of the element are not the parsed character data, if the end tag corresponding to the start tag defined appears, terminating, and if the end tag does not appear, ignoring and returning.
7. The method of claim 6, wherein the step (c) comprises the steps of:
- 15 if the read token does not include a defined start tag, reading the data continuously until the end tag appears, thereby ignoring the token; and
- reading a new token.
8. A recording medium for storing a program for parsing a called web-
- 20 document of a web-server, the recording medium being read by a computer, the program comprising the functions of:
- (a) reading a token from the web-document and parsing the token;
- (b) if the token is not a defined start tag or if the token is a comment or a space as result of the function (a), ignoring the token, and when the defined start tag is read, parsing
- 25 an attribute of an element from the token;
- (c) parsing the attribute of the element from the token, storing GUI-related information of the element, and parsing contents of the element;
- (d) if the contents of the element are parsed character data as result of the function (c), storing GUI-related information of the contents, and if the contents of the element are
- 30 not the parsed character data, reading data until an end tag appears; and
- (e) in case the contents of the element are not the parsed character data, if the end tag corresponding to the start tag defined appears, terminating, and if the end tag does not appear, ignoring and returning.
9. A system for parsing a web-document based on elements to provide contents thereof to a handheld terminal, comprising:
- 35

PCT/KR2003/002569

a syntax parser for parsing syntax for the tokens extracted and separated by the word parser on the basis of contents model, and generating an object on the basis of GUI of the terminal.

11. The system of claim 9, wherein the word parser comprises:
a comment parser for recognizing a comment or a space and generating a token;
a markup start parser for recognizing a markup start tag and generating a token;
an attribute parser for parsing an attribute and generating a token; and
a parsed character data analyzer for analyzing parsed character data and generating
a token.

13. The system of claim 9, wherein the contents model means a hierarchy of elements and an attribute list, and is defined in DTD for all documents provided to the handheld terminal.

19

WO 2004/049194

PCT/KR2003/002569

15. A system for parsing web-document based on elements, comprising:
a token table comprising tokens defined in an XML document, keywords defined in document type definition (DTD) for documents provided to a handheld terminal, and a list of elements, which can be supported by each terminal;
5 a word parser for extracting and separating tokens of the web-document supplied to the terminal regardless of kind of a markup language used to compose the web-document by referring to the token table;
a contents model determined by DTDs for the documents provided to the terminal that includes a hierarchy of elements and an attribute list; and
10 a syntax parser for parsing syntax for the tokens extracted and separated by the word parser on the basis of the contents model, and generating an object on the basis of GUI of the terminal through the parsed syntax.

16. The system of claim 15, the word parser comprises:
15 a comment parser for recognizing a comment or a space and generating a token;
a markup start parser for recognizing a markup start tag and generating a token;
an attribute parser for parsing an attribute and generating a token; and
a parsed character data analyzer for analyzing parsed character data and generating
a token.

- 20 17. The system of claim 15, wherein the word parser comprises a token generator and an XML well-formedness verifier, receives the supplied document character by character, recognizes a token of the document on the basis of the token table, and extracts the token by using the token generator and the XML well-formedness verifier.

- 25 18. The system of claim 15, wherein the syntax parser comprises:
an XML verifier for verifying whether a supplied web-document is composed suitable for each DTD supplied on the basis of the token extracted and separated by the word parser; and
30 a GUI-based object generator for matching the parsed syntax and a GUI of the terminal.

19. A handheld terminal comprising:
an integral parser for parsing a web-document composed of a predetermined
35 markup language supplied from a web-server;
a memory for storing information parsed by the integral parser; and

WO 2004/049194

PCT/KR2003/002569

an application program using information extracted from the integral parser.

20. A handheld terminal comprising an antenna, a CPU, a peripheral circuit, a vocoder, a memory and an audio codec, wherein the memory comprising:

5 an integral parser for calling a web-document supplied from a web-server regardless of a markup language used to compose the web-document and parsing the web-document on the basis of elements; and

an application program using information extracted from the integral parser.

10 21. The handheld terminal of claim 19 or 20, wherein the integral parser comprises:

a token table comprising tokens defined in an XML document, keywords defined in DTD for all documents provided to the handheld terminal, and a list of elements, which can be supported by each of the handheld terminals;

15 a word parser for extracting and separating all tokens of the document supplied to the terminal regardless of kind of a markup language used to compose the web-document by referring to a token table;

a contents model defined in DTD for all documents provided to the terminal and meaning a hierarchy of the elements and an attribute list; and

20 a syntax parser for parsing syntax for the tokens extracted and separated by the word parser on the basis of contents model, and generating an object on the basis of GUI of the terminal through the parsed syntax.

22. The system of claim 21, the word parser comprises:

25 a comment parser for recognizing a comment or a space and generating a token;

a markup start parser for recognizing a markup start tag and generating a token;

an attribute parser for parsing an attribute and generating a token; and

a parsed character data analyzer for analyzing parsed character data and generating a token.

30

23. The system of claim 21, wherein the word parser comprises a token generator and an XML well-formedness verifier, receives the supplied document character by character, recognizes a token of the document on the basis of the token table, and extracts the token by using the token generator and the XML well-formedness verifier.

35

24. The system of claim 21, wherein the syntax parser comprises:

WO 2004/049194

PCT/KR2003/002569

an XML verifier for verifying whether a supplied web-document is composed suitable for each DTD supplied on the basis of the token extracted and separated by the word parser; and

5 a GUI-based object generator for matching the parsed syntax and a GUI of the terminal.

25. The handheld terminal of claim 19 or 20, wherein the application program comprises an object based on a GUI of the handheld terminal.

10 26. A method for parsing a web-document supplied from a web-server, the web-document being composed of a predetermined markup language, the method comprising the steps of:

(a) reading a token from the web-document by referring to a token table, extracting and separating the token;

15 (b) if the extracted and separated token is not a defined start tag or if the token is a comment or a space, ignoring the token;

(c) when the extracted and separated token is recognized as the defined start tag, parsing an attribute of an element from the token and storing GUI-related information of the element;

20 (d) parsing contents of the element after parsing the attribute of the element;

(e) as the result of the step (d), if the contents of the element are parsed character data, storing GUI-related information of the contents, and if the contents of the element are not the parsed character data, determining whether an end tag appears;

25 (f) as the result of the step (e), if the end tag does not appear, repeating from the step (a), and if the end tag appears, determining whether the end tag corresponds to the defined start tag; and

(h) as the result of the step (f), if the end tag corresponds to the defined start tag, terminating, and otherwise, ignoring and returning.

30 27. The method of claim 26, wherein the step (c) comprises the steps of:
if the extracted and separated token does not include a defined start tag, reading the data continuously until the end tag appears, thereby ignoring the token; and
reading a new token.

35

WO 2004/049194

PCT/KR2003/002569

28. A handheld terminal, comprising:

an integral parser for parsing web-documents composed of a plurality of predetermined markup languages on the basis of elements;

a memory for storing information parsed by the integral parser; and

5 an application program using information extracted from the integral parser.

29. A system, comprising:

a content provider configured to provide first type documents using a first markup language and second type documents using a second markup language different from the first markup language; and

10 a handheld terminal that receives the first and second type documents from the content provider, wherein the handheld terminal comprises,

an integral parser configured to parse both a first type document and a second type document on the basis of elements to extract information thereof, and

15 an application program configured to receive the information extracted from the integral parser.